Serial Correlation in Time Series Regression Models – an introduction

Artur Silva Lopes

Version 1.03, September 30, 2021

ISEG, ULisboa

Preface

Although the expression "serial correlation" (or simply autocorrelation) may be used in relation to any variable, in traditional econometrics it is meant to refer to a particular variable or time series: that of the errors of the regression model. It has therefore a clearly negative sense in this framework: the errors of a well specified model should not be serially correlated (unless in very special circumstances). Although I have tried always to add the name of the variable to which it refers, any omission should be interpreted in the usual sense.

This text is very strongly based in my teaching notes, most of them with many years as the subject has not evolved or has changed little in the last years. Anyway, I had to make a choice between many teaching materials and I chose not to address the problems of GLS and EGLS estimation. It is not that they are completely irrelevant. For instance, some of the most powerful unit root tests require GLS estimation. However, as will become clear below, I think that they are very rarely a good option to follow when one faces serial correlation problems and, given severe time restrictions, I have opted to neglect the theme. Although attractive from the computational point of view, they are only rarely a reasonable alternative to folllow. As Professor Grayham Mizon noted: "*autocorrelation correction: Don't!*" My hope is that the reason why such a strategy is usually flawed is clearly understood with this text.

With a very "classical" background but philosophy-based in the British (LSE) econometric tradition, this text draws heavily from such classical sources as those of Davidson and Mackinnon (1993), Greene (2012), Johnston and Dinardo (1998). With the inclusion of some examples of Monte Carlo simulation it is expected that some testing problems are well grasped.

March 13, 2020

. .

.

1 Introduction

In the time series linear regression model $y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t, \;\; t = 1, 2, \ldots, T$, i. e., in matrix form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, with $\mathrm{E}(\mathbf{u}|\mathbf{X}) = \mathbf{0}$, the hypothesis that

$$\mathrm{Var}(\mathbf{u}|\mathbf{X}) = \mathrm{E}(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \sigma^2\mathbf{I},$$

may be violated, that is, one may face

$$\mathrm{Var}(\mathbf{u}|\mathbf{X}) = \mathrm{E}(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \boldsymbol{\Omega} = \sigma^2\mathbf{V} \neq \sigma^2\mathbf{I},$$

due to serial correlation of the errors:

$$\exists t \neq s : \mathrm{Cov}(u_t, u_s) = \mathrm{E}(u_t u_s) \neq 0.$$

Strict exogeneity of the regressors was assumed above to facilitate the representation in matrix form but this hypothesis is not crucial to present the problem. Actually, error serial correlation is often presented only unconditionally, as in the third equation above, and hence it appears that the nature of the regressors is irrelevant. However, as we will see below, there are many instances where this nature is crucial.

In case the errors are serially correlated, assuming nevertheless that they are stationary (and therefore homoskedastic), their covariance matrix may be written as

$$\boldsymbol{\Omega} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{T-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{t-3} & \cdots & \gamma_0 \end{bmatrix} = \gamma_0 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{T-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{t-3} & \cdots & 1 \end{bmatrix},$$

where γ_0 denotes the (assumed common) error variance, the parameter usually represented with σ^2. The second matrix is the $\mathbf{V}$ matrix, $\mathbf{V} = [v_{ts}] = \left[\frac{\gamma_{|t-s|}}{\gamma_0}\right] = [\rho_{|t-s|}]$, that is, it is the matrix of the serial correlation coefficients.

2 Some sources of "residual autocorrelation"

As usual, since the errors of the regression model are not observed, to assess whether they are serially correlated one needs to replace them by the OLS residuals,

e_t (as in the analysis of homoskedasticity). Therefore, sometimes we may say that we get symptoms of "residual autocorrelation", though we are really referring to the autocorrelation of the errors. What may originate these "symptoms of residual autocorrelation"?

In some econometric models and mostly many years ago, but also in some time series literature, error autocorrelation was employed as a normal or natural way to provide some dynamics to a static model or to augment the dynamic effects of the model. Thus, in a sense, it was even welcome. A more modern econometric perspective is radically different: a correctly specified model must not have serially correlated errors. All that is systematic, that has some pattern, must be present in the main or principal component of the model, in the "signal", it must not be left to the "noise" component.

Therefore, OLS residuals should reflect this property (at least partially). Most likely, in case they appear to be autocorrelated, that should be taken as a sign of model misspecification, of a poor dynamic specification. In particular, it is now well known that symptoms of (error) serial correlation provided by some test statistic are frequently due to:

a) misspecification errors in the adopted functional form, e.g., using a linear function when a non-linear one, concave or convex, should be considered instead;

b) omitted regressors and particularly an insufficient dynamic specification.

Note that point a) may be easily graphically illustrated. Some further material concerning b) will be addressed later.

This should made clear that residual autocorrelation should not be automatically interpreted as a problem that is intrinsic to the errors. Usually there is no "evil" inherent to them. Instead, most often it is only a visible reflection of an incorrect specification of the regression function.

3 The case of strictly exogenous regressors

Let us assume that the regressors are strictly exogenous, that is, that $\mathrm{E}(u_t|\mathbf{X}) = 0, \forall t$, implying that $\mathrm{Cov}(x_{tj}, u_s) = 0, \forall t, s, j$.

In this case, the only purpose of the econometric analysis is usually one of testing a certain statistical hypothesis with a very specific economic meaning. In other

words, the model is not specified to analyse any dynamic effects, or to provide a good empirical (usually dynamic) description of the relations between the variables or to provide sound forecasts, etc. Still put differently, usually one is not free to specify the model according to any of these purposes. Testing must be made in the framework of a tight specification that cannot be changed to adapt to the data. There is no freedom to bring the model close to the data and its adherence is considered irrelevant. Examples of this framework are testing the rational expectations hypothesis or testing the efficient market hypothesis in the context of certain (usually static) equations.

What are the consequences of error autocorrelation for OLS estimation and inference in this case?

1. **Unbiasedness.** It is easy to show that the OLS estimator is still unbiased, $\mathrm{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, because strict exogeneity is not affected ($\mathrm{E}(\mathbf{u}|\mathbf{X}) = \mathbf{0}$).

2. **Covariance matrix.** However, it is easy to show that the covariance matrix of $\widehat{\boldsymbol{\beta}}$ is no longer the usual $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$:

$$\begin{aligned}
\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\mathbf{X}) &= \mathrm{E}[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\mathbf{X}] \\
&= \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|\mathbf{X}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}(\mathbf{u}\mathbf{u}'|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}$$

3. **Inference.** Since $\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ was the basis for the inference methods associated with OLS, these methods are no longer valid (not even asymptotically). Hence, when using the critical values taken from the t and F distributions (for the t and F statistics, respectively), we can be making incorrect inferences with a (much) larger probability than we think, possibly (much) more frequently than is assumed.

   As a simple theoretical example, consider the case presented in Wooldridge (2016, p. 373) of the simple regression model $y_t = \beta_1 + \beta_2 x_t + u_t$, with the errors following a stationary AR(1) process, $u_t = \rho u_{t-1} + \epsilon_t$, $|\rho| < 1$, $\epsilon_t \sim iid(0, \sigma^2)$. For convenience, let us assume that $\bar{x} = 0$. Then, it can be shown that

$$\mathrm{Var}(\widehat{\beta}_2|\mathbf{X}) = \frac{\sigma_u^2}{TSS_x} + 2\frac{\sigma_u^2}{TSS_x^2}\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}\rho^j x_t x_{t+j},$$

where TSS_x denotes the variation of the regressor, $\sum x_t^2$. Hence, the first term of the right-hand side (RHS) denotes the usual expression for $\text{Var}(\widehat{\beta}_2|\mathbf{X})$ when there is no serial error correlation, that is, when $\rho = 0$ (and hence $u_t \equiv \epsilon_t$). Since the usual estimator for $\text{Var}(\widehat{\beta}_2|\mathbf{X})$ neglects the second term in the RHS and since the most frequent situation is one of positive serial correlation for both the error term and the regressor, usually it tends to underestimate the true variance. This is because (besides $TSS_x > 0$ and $\sigma_u^2 > 0$, obviously) usually $\rho > 0$ and most terms $x_t x_{t+j}$ are also positive[1]. When ρ is large and the serial correlation of x_t is also large (and positive), this underestimation can be very strong, that is, we may be led to consider that the OLS estimator of β_2 is much more precise than it really is.

As a consequence, when testing hypothesis about β_2, the null hypothesis will be rejected (much) more often than it should be when the null is true, i.e., (much) more frequently than the predetermined nominal size of the test. This is particularly easy to understand for the case of t-statistics because their denominators are the deflated standard errors. The real size of the tests will then become (much) larger than the nominal size, that is, over-rejections of the true null hypothesis will tend to occur. In some cases these over-rejections can be very severe. Sometimes they are also called spurious rejections because the true null hypothesis is erroneously rejected (much) more often than it should (the usual 5 or 10% of the times).

Monte Carlo illustration. To illustrate the previous example consider that the DGP is

$$y_t = 2 + 1\, x_t + u_t, \quad u_t = \rho u_{t-1} + \epsilon_{1t}, \ |\rho| < 1,$$

with

$$x_t = \lambda x_{t-1} + \epsilon_{2t}, \quad |\lambda| < 1,$$

and

$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \sim iid\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

Hence, both the error term and the regressor follow a stationary AR(1) process. Suppose that we are interested in testing the (true null)

$$H_0 : \beta_2 = 1, \quad vs. \quad H_1 : \beta_2 \neq 1,$$

[1] Since $\bar{x} = 0$ these are the terms of the sum in the numerator of the empirical autocovariances.

and that the sample size is $T = 50$. The following table collects the results of the Monte Carlo experiments with $20,000$ replications presenting the percentage times that H_0 was incorrectly rejected.

Real size estimates (in %) for nominal 5% tests with $\lambda = 0.9$

ρ	0.0	0.2	0.4	0.6	0.8	1.0
% reje.	5.40	10.48	17.60	26.27	38.92	54.43

We can observe that when there is no serial correlation of the errors the estimated real size (5.4%) is very close to the nominal one but that as ρ grows the rejection frequencies tend to rise very fast and, in the extreme case where the errors contain a unit root (and hence are non stationary) the rejections dramatically exceed 50%: a true null hypothesis is rejected more often than not. This case may be considered as a caricature because it exceeds the stationary context but it illustrates many real situations. Even when $\rho = 0.2$ only, which represents weak autocorrelation, over-rejections are already significant, at about 10.5%, clearly exceeding the nominal size. And as mentioned, as ρ increases the problem clearly worsens attaining a real size estimate of 54.43% at the extreme case when $\rho = 1$, that is, when there is a unit root.

The problem is not one of a small sample size. Indeed, if we increase the sample size the problem will tend to get worse, with a larger proportion of spurious rejections. Moreover, notice also that as this example is typical of many cases in macroeconometrics, the problem with serial correlation is more clearly defined than the similar problem with heteroskedasticity: it usually implies overestimation of the precision of the OLS estimator and consequent over-rejection of true null hypothesis.

Notice also that the well known case of spurious regressions is a simple example of this problem, i.e., it is simply a special case of this one. This is the most serious consequence of these size distortions or over-rejections ($\alpha_{real} > \alpha_{nominal}$), as they are usually called: the "discovery" of significant, seemingly causal relations, that do not really exist. However, we leave this case to further study ahead, when unit root testing is addressed.

4. Consistency. If

$$\text{plim}\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right) = \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right) = \mathbf{\Sigma}_{xx},$$

$$\text{plim}\left(\frac{1}{T}\mathbf{X}'\mathbf{u}\right) = \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t u_t\right) = \text{plim}(\bar{\mathbf{g}}) = \text{E}(\mathbf{g}_t) = 0,$$

and provided the following condition is also satisfied

$$\text{plim}\left(\frac{1}{T}\mathbf{X}'\mathbf{V}\mathbf{X}\right) = \mathbf{Q}^*,$$

a positive definite matrix, it can be shown that the OLS estimator of the coefficients is still consistent (Greene, 2012), that is $\text{plim}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. The first condition is satisfied when the regressors are stationary (and ergodic), so that the sample second moments converge in probability to the corresponding population moments. The second condition also demands stationarity of both regressors and errors besides contemporaneous orthogonality. Note however that the process $\mathbf{g}_t \equiv \mathbf{x}_t u_t$ now cannot be a martingale difference sequence due to the serial correlation of the errors. Nevertheless, provided $\text{E}(\mathbf{g}_t) = \mathbf{0}, \forall t$, the stationarity of both $\mathbf{x}_t$ and u_t is sufficient to warrant convergence to 0 of the sample moment $\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t u_t$ through the ergodic theorem. According to Greene (2012), the last condition is satisfied when (again) both regressors and errors are stationary and ergodic. For instance, in the case that u_t follows an AR(1) process, its stationarity ($|\rho| < 1$) is sufficient to guarantee its ergodicity as well.

5. **Asymptotic normality.** To derive the asymptotic normality of the OLS estimator, as usual it is necessary to show that

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{X}'\mathbf{u} \tag{1}$$

is asymptotically normally distributed. Now the problem is that applying a central limit theorem to serially uncorrelated variables is not possible because

$$\frac{1}{\sqrt{T}}\mathbf{X}'\mathbf{u} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{u}_t$$

is not a sum of variables complying with that condition. Instead, they are obviously serially correlated.

There is, however, a central limit theorem that, again in the context of stationary (and ergodic) regressors and errors, warrants asymptotic normality of this

sum. This theorem appears in Hayashi (2000) under the heading of Gordin's theorem and requires more time to grasp than we have available[2].

Hence, with the exception of some extreme cases, it is possible to show that generally

$$\widehat{\boldsymbol{\beta}}|\mathbf{X} \overset{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}].$$

6. **Efficiency.** Since the assumption of no serial correlation is essential to demonstrate the Gauss-Markov theorem, its violation implies that the OLS estimator is no longer best among the linear unbiased estimators of $\boldsymbol{\beta}$, i. e., that it is no longer BLUE.

   When the $\mathbf{V}$ matrix is known, it is the GLS estimator that is BLUE. However, the assumption of a known $\mathbf{V}$ matrix is highly non-realistic, except for a few cases where circumstances are very special. Therefore, usually GLS is unfeasible, and a FGLS (feasible GLS) or EGLS (estimated GLS) must be employed. Under general conditions this is asymptotically efficient. However, in many cases its asymptotic superiority really requires large samples to be observed, that is, even when there really exists error autocorrelation, in some cases the OLS estimator may be more efficient than its FGLS competitor.

7. **Robust inference. HAC.** Provided that consistency of OLS estimation is assured, it can be used to provide asymptotically valid inferences, based on an also consistent estimator of its covariance matrix. Again, recall that this is usually the case when an accurate dynamic description is not necessary and therefore efficiency in estimation is also unimportant.

   In the case where the only problem is heteroskedasticity, estimation of the covariance matrix is based in

$$\widehat{\mathbf{S}}_0 = \frac{1}{T}\sum_{t=1}^{T} e_t^2 \mathbf{x}_t \mathbf{x}_t',$$

   the well known matrix entering the White estimator, because the off diagonal elements of the error covariance matrix would be all equal to zero. Now, however, serial correlation implies that those elements are non null and therefore the matrix becomes rather more complicated. Moreover, it is necessary that

[2] The first recommended reading on this subject is Hayashi (2000), section 6.5.

the matrix is at least semi-definite positive. With this in mind, Newey and West (1987) proposed the following nonparametric estimator, which became thereafter known as the Newey-West estimator or HAC (heteroskedastic and autocorrelation consistent) estimator of the OLS covariance estimator:

$$\widehat{\mathbf{S}}^* = \frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\widehat{\mathbf{S}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1},$$

where

$$\widehat{\mathbf{S}} = \widehat{\mathbf{\Omega}}_0 + \sum_{j=1}^{L}\omega_j(\widehat{\mathbf{\Omega}}_j + \widehat{\mathbf{\Omega}}_j'),$$

with

$$\widehat{\Omega}_j = \sum_{t=j+1}^{T} e_t e_{t-j}(\mathbf{x}_t\mathbf{x}_{t-j}')$$

and the weights ω_j are usually given by $\omega_j = 1 - \frac{j}{L+1}$, $j = 1, \ldots, L$, which are known as the Bartlett weights.

The parameter L, sometimes called the lag truncation parameter, denotes the order until which the autocorrelations of the errors (residuals) must be considered, that is, the order $L + 1$ is the first one that can be considered negligible (when $L = 0$ one gets the usual White estimator). Sometimes L is also called the bandwidth. In case the errors follow a MA process, that order is usually low because, as is well known, a MA process has short memory, that is, all autocorrelations with order higher than its order are null. However, for an AR process, even of a low order, the autocorrelations are never equal to zero, though they approach zero. Therefore, without specifying a model for the autocorrelation, this parameter is hard to determine and its importance may, sometimes, be large. A fast rule, mentioned in Stock and Watson (2015), is to make $L = 0.75\,T^{1/3}$. However, several different values of L should be tried in search for robustness. In case the results are very dependent on the value of L a finer analysis is necessary. For instance, it is not unusual to find the recommendation to fix $L = \max\{p, q\}$, p and q denoting the higher order of the autocorrelation coefficients and the partial correlation coefficients of the residuals that is significant, respectively.

The weights ω_j are needed to ensure that the estimator is a semi-definite matrix (otherwise we can get negative estimates for some linear combinations

of the variances) and sometimes they are called "the kernel". The expressions above define the Bartlett kernel, the most popular. Two other well known kernels are the Parzen kernel and the quadratic spectral kernel (see, e. g., Hayashi, 2000, pp. 411-12)[3].

Monte Carlo illustration. To illustrate the use of the NW estimator and the corresponding inference let us consider again the same DGP of the previous example though with a more harmless (albeit less realistic) parameter value controlling the serial correlation of the regressor, $\lambda = 0.7$. Consider testing the (true) null hypothesis $H_0 : \beta_1 = 1$ vs. $H_1 : \beta_2 \neq 1$ again, using three alternative statistics:

- the usual OLS t-statistic, as in the previous example (t_{OLS});
- the t-statistic built with a standard error taken from the the Newey-West matrix with $L = 5$ and the Bartlett kernel (t_{NW-5});
- and the t-statistic built with a standard error taken from the Newey-West matrix using the Bartlett kernel as well but with $L = 10$ (t_{NW-10}).

The sample size is moderately large, with $T = 100$, and the main purpose of the study is to assess the quality of inferences provided by the NW estimator, i.e., to assess whether the large sample distribution using the NW estimator provides a good approximation to the (unknown) small sample distribution. The following table presents the size estimates for several values of the parameter ρ, obtained with 10000 replications.

Real size estimates (in %) for nominal 5% tests with $\lambda = 0.7$

ρ	0.0	0.2	0.4	0.6	0.8	1.0
t_{OLS}	4.87	8.76	14.39	21.76	30.67	38.84
t_{NW-5}	5.99	6.90	8.27	9.98	12.68	15.53
t_{NW-10}	6.25	6.90	7.75	8.90	10.86	12.64

The following conclusions emerge:

a) the size distortion (or over- or spurious rejection) problem becomes serious for OLS when $\rho > 0.2$ but it does not get as serious as in the previous case because the regressor is less autocorrelated;

[3] This estimation is available in TSP, as a GMM option, when the instruments are the same as the regressors: `GMM (HET(NOHET), NMA=L, INST=(same as regressors)|) ....`

b) although both NW-based t-statistics alleviate the problem they never remove it completely (and obviously they are not designed to handle non-stationarity);

c) even when ρ is only 0.8 and $L = 10$ the NW method is still affected by significant size distortion (though it performs a bit better than when $L = 5$);

d) a larger sample size is likely needed for the NW method to perform better and this could be used to (implicitly) estimate higher order autocorrelation coefficients (and to estimate more efficiently the lower order ones).

e) when ρ is small there is no gain in considering a higher ($L = 10$) truncation parameter. However, it suffices that $\rho = 0.4$ to get a better size performance with the test based in $L = 10$ than with $L = 5$.

To work reasonably well the method does indeed require large sample sizes. To understand this better we must get back to the second term in the right hand side of equation (1), $\frac{1}{\sqrt{T}}\sum \mathbf{x}_t u_t = \frac{1}{\sqrt{T}}\sum \mathbf{g}_t$. When the $\mathbf{g}_t$ process is stationary and weakly dependent and under additional ("Gordin's") conditions, a specialized central limit theorem (or "mixing" CLT) implies that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{g}_t \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S}),$$

where $\mathbf{S} = \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j$, with $\mathbf{\Gamma}_j = \mathrm{E}[(\mathbf{g}_t - \boldsymbol{\mu})(\mathbf{g}_{t-j} - \boldsymbol{\mu})'] = \mathrm{E}(\mathbf{g}_t \mathbf{g}'_{t-j})$, is called the long-run covariance matrix of the $\mathbf{g}_t$ process. Since

$$\mathbf{S} = \lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{g}_t\right) = \lim_{T\to\infty} \mathrm{Var}(\sqrt{T}\bar{\mathbf{g}}),$$

to understand the problems involved in its estimation let us consider the simple case where the g_t process is scalar, that is, when there is only one regressor.

$$\begin{aligned}
\mathrm{Var}(\tfrac{1}{\sqrt{T}}\textstyle\sum_{t=1}^{T} g_t) &= \tfrac{1}{T}\mathrm{Var}(g_1 + g_2 + \ldots + g_T) \\
&= \tfrac{1}{T}[\mathrm{Cov}(g_1, g_1 + ... + g_T) + \mathrm{Cov}(g_2, g_1 + g_2 + \ldots g_T) + \ldots + \\
&\quad \mathrm{Cov}(g_T, g_1 + g_2 + \ldots + g_T)] \\
&= \tfrac{1}{T}[(\gamma_0 + \gamma_1 + \ldots + \gamma_{T-1}) + (\gamma_1 + \gamma_0 + \gamma_1 + \ldots + \gamma_{T-2}) + \ldots + \\
&\quad (\gamma_{T-1} + \gamma_{T-2} + \ldots + \gamma_1 + \gamma_0)] \\
&= \tfrac{1}{T}[T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \ldots + 2\gamma_{T-1}] \\
&= \gamma_0 + 2\textstyle\sum_{j=1}^{T-1}(1 - \frac{j}{T})\gamma_j.
\end{aligned}$$

Therefore[4],

$$\lim_{T\to\infty} \text{Var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t) = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j = \sum_{j=-\infty}^{\infty}\gamma_j.$$

The problem becomes quite clear: how to estimate consistently and efficiently a magnitude that depends on a non finite number of parameters, the autocovariances of the g_t process? On the one hand, estimating only a few of them means that serial correlation that might exist at higher lags is being neglected, thereby leading to an inconsistent estimator. On the other hand, trying to estimate a large number of autocovariances implies that the number of observations used to estimate those of higher order may be very low, resulting in an inefficient estimator. Moreover, the estimator must be (semi-definite) positive. A balance is achieved making the truncation depend on the sample size. But this may be insufficient to obtain a good estimate and hence to obtain tests with a good size performance.

Notice, however, that although a great effort has been devoted to non-parametric estimators related to the one by Newey and West, simpler, parametric estimators also exist that may perform reasonably well in small samples. See the appendix and e. g., Hamilton (1994). In any case, our expectations on this topic should be low: as West (2010) emphasizes, "*the performance of all estimators leaves much to be desired*".

Empirical example. To illustrate empirically the use of HAC estimation let us return to the empirical example of the first chapter, the case of private consumption in the portuguese economy. Recall that the prefix D in all the variables' names represents Δ, the first difference operator; LC, LR and LS are, respectively, the logs of consumption, income, and of a wage index. INF denotes inflation.

The first equation below recalls OLS estimation, now in the company of some test statistics to detect serial correlation. As will become clear in the last section, there is no sign of the problem, the two main statistics, BG(1) and BG(3), with p-values largely exceeding the usual 5% or 10%, i.e., the null hypothesis of no serial correlation clearly supported.

[4]For those familiar with spectral analysis, recall that the spectra of a stationary process at frequency ω is $S(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_j e^{-i\omega j}$. Hence, $S = 2\pi \times S(0)$.

```
                                    Equation   1
                                    ============
                     Method of estimation = Ordinary Least Squares
Dependent variable: DLC
Current sample:  1966 to 1995
Number of observations:  30
  Std. error of regression = .018050
                R-squared = .652677
            Durbin-Watson = 1.73048 [.079,.454]
Breusch/Godfrey LM: AR/MA1 = .508995 [.476]
Breusch/Godfrey LM: AR/MA3 = 2.75500 [.431]

           Estimated    Standard
Variable  Coefficient     Error       t-statistic   P-value
C         .018740       .460240E-02   4.07185       [.000]
DLR       .330562       .100103       3.30221       [.003]
DLS       .375609       .085867       4.37429       [.000]
DINF      .294196E-02   .853164E-03   3.44830       [.002]

                        GENERALIZED METHOD OF MOMENTS
                        ============================
                          WITH STARTING VALUES VIA:
                       NONLINEAR TWO STAGE LEAST SQUARES
EQUATIONS: EQ
INSTRUMENTS: C DLR DLS DINF
CONVERGENCE ACHIEVED AFTER   1 ITERATIONS
 Number of observations = 30  E'PZ*E = .122894E-31
                         Standard
Parameter  Estimate        Error       t-statistic   P-value
BEQ0       .018740       .447534E-02   4.18745       [.000]
BEQ1       .330562       .096896       3.41151       [.001]
BEQ2       .375609       .084745       4.43221       [.000]
BEQ3       .294196E-02   .793705E-03   3.70662       [.000]
 Standard Errors computed from quadratic form of analytic first derivatives(Gauss
(also robust to autocorrelation:  NMA= 4,  Kernel=Bartlett)
```

Therefore, there is no support to use HAC estimation: previously we have noticed that correcting the standard errors for heteroskedasticity appeared unnecessary and now a similar case occurs in relation to serial correlation. Nevertheless, just to illustrate the estimation of the covariance matrix with the Newey-West method, the second set of results exemplifies that estimation. Notice, however, that much of the original TSP output was deleted as it was irrelevant. The bandwidth parameter was set equal to 4, which is a small but adequate value in this case, justified both by the small sample size and the absence of autocorrelation symptons. The `NOHET` option was chosen, i.e., robustness is searched only with respect to autocorrelation. The estimated parameters are denoted with `BEQ0` to `BEQ3` and they are obviously the same as before. The corrected standard errors differ so slightly from the original that the robust t-statistics are also practically the same (and this is also an additional sign that serial correlation must be negligible).

This is not a typical example. Very often the evidence for positive first order autocorrelation is strong, and in that case a larger bandwidth parameter would be more appropriate, even with such a small sample size. Often, the robust standard errors are much larger than the (OLS) originals, and a finding for significance is reverted, the robust t-statistic insufficiently large (in absolute value) to allow rejecting the null hypothesis.

4 The case of dynamic models

Following the previous logic, this section should be called "the case of predetermined regressors" or similar. There are, however, two problems with such a title: a) as we will see shortly, the presence of autocorrelated errors changes the nature of some regressors, that cease to be predetermined and become endogenous; b) the main purpose of this section really is to address the case of dynamic models, when they are designed mostly to provide an accurate description of inter-temporal relations between variables and the nature of the regressors is relatively unimportant. Moreover, since a dynamic model usually requires at least one lag of the dependent variable to provide a good approximation to the data, the presence of (at least) y_{t-1} in the right hand side of the equation is implicitly assumed in the dynamic model.

The framework is now rather different from the one of the previous section: now we have the freedom to specify the model according to our convenience; we are not restricted to work with a model that is usually static. Instead, the purpose is to

build a good model to:

a) provide an accurate description of the economy (or, more frequently and modestly, of a portion of it);

b) to make reliable forecasts;

c) to analyse dynamic inter-relationships between variables so that, for instance, a sound economic policy may be designed to attain certain effects at certain times, etc. .

To begin, recall the classical cases of endogenous regressors:

i) omitted regressors correlated with those that are included and an incorrect functional form;

ii) measurement errors in the regressors, when they are correlated with the model errors (as in the classical error in variables case);

iii) endogeneity due to a simultaneous determination of variables (that is, a contemporaneous feedback effect).

A new case is now added to these: certain types of serially correlated errors in certain dynamic models and, particularly in those where there is at least a lagged dependent variable as regressor. In all these cases,

$$\mathrm{E}(\mathbf{g}_t) = \mathrm{E}(\mathbf{x}_t u_t) \neq 0 \Rightarrow \mathrm{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t u_t\right) = \mathrm{plim}\left(\frac{1}{T}\mathbf{X}'\mathbf{u}\right) \neq \mathbf{0} \Rightarrow \mathrm{plim}\widehat{\boldsymbol{\beta}} \neq \boldsymbol{\beta},$$

that is, since the orthogonality condition fails, the OLS estimator of the regression coefficients (besides biased) is not (even) consistent.

In what follows a very simple example will be presented. We know that in the stationary AR(1) model

$$y_t = \beta y_{t-1} + u_t,\ |\beta| < 1,\ \text{with}\ \ u_t \equiv \epsilon_t \sim iid(0, \sigma^2),$$

the regressor y_{t-1} is pre-determined and that, though biased (to underestimation), the OLS estimator of β is consistent. Suppose now that we have instead

$$u_t = \rho u_{t-1} + \epsilon_t,\ |\rho| < 1,$$

i.e., the errors follow a stationary AR(1) process and therefore the model for y_t is no longer a strict AR(1). This seemingly small change will affect the nature of the regressor y_{t-1}. Indeed,

$$\begin{aligned}\text{Cov}(y_{t-1}, u_t) &= \text{Cov}(\beta y_{t-2} + u_{t-1}, \rho u_{t-1} + \epsilon_t) \\ &= \beta\rho\text{Cov}(y_{t-2}, u_{t-1}) + \rho\sigma_u^2 \\ &= \beta\rho\text{Cov}(y_{t-1} + u_t) + \rho\sigma_u^2,\end{aligned}$$

where the last equality derives from the stationary (and ergodic) environment where we are operating. Therefore,

$$\text{Cov}(y_{t-1}, u_t) = \frac{\rho\sigma_u^2}{1-\beta\rho} \neq 0 \text{ (if} \rho \neq 0).$$

Hence,

$$\text{E}(y_{t-1}, u_t) \neq 0 \Rightarrow \text{plim}(\widehat{\boldsymbol{\beta}}) \neq \boldsymbol{\beta},$$

that is, the OLS estimator becomes inconsistent because the regressor is now endogenous. Therefore, inference now becomes impossible; it simply cannot be based on an inconsistent estimator.

In conclusion, in certain models the status of lagged dependent variables depends on the serial correlation properties of the errors. This is a classical case of regressor endogeneity that led many authors to claim, somewhat too rapidly, that the mixture of these two situations always conducted to an inconsistency problem of the OLS estimator. As Wooldridge (2016) argues, that may be not always the case, that is, there are cases where the two situations might live peacefully with each other, without violating the orthogonality condition (and therefore OLS may remain consistent in those cases). However, it is true that in the most current cases the mixture is dangerous. For instance, replacing the assumption of a stationary AR(1) process for the errors with one of a MA(1) process changes only a few details in the example, OLS becoming inconsistent again.

Given the mentioned purpose, the specification search is, in general, for a dynamically complete model. Now, the model $y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t$ (where x_t' may contain lagged variables in spite of the subscript) is said to be a dynamically complete model if

$$\text{E}(y_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots) = \text{E}(y_t|\mathbf{x}_t),$$

that is, if the additional information, the information that was previously absent from the initial model is irrelevant; given the information that the model already contains (in $\mathbf{x}_t$), it does not add anything useful; it doesn't help to explain the behaviour of y_t besides the explanation that is already provided by $\mathbf{x_t}$. In other words, the dynamic explanation provided by $\mathbf{x_t}$ is sufficient; further historical information is useless.

It can be shown (see Wooldridge, 2016, p. 362)[5] that if a model is dynamically complete, then its errors comply with the condition

$$\mathrm{E}(u_t|\mathbf{x}_t, u_{t-1}, \mathbf{x}_{t-1}, u_{t-2}, \ldots) = 0, \forall t,$$

which is precisely the sufficient condition that was presented earlier for the $\mathbf{g}_t$ process ($\mathbf{g}_t = \mathbf{x}_t u_t$) to be a martingale difference process (m.d.s.) (recall assumption H5' of the model for predetermined regressors). From this, it follows that when a model is dynamically complete its errors are serially uncorrelated:

$$\mathrm{E}(u_t|u_{t-1}, u_{t-2}, \ldots) = 0, \forall t.$$

Hence, when we find symptoms of residual autocorrelation, we may interpret them as signs or as indirect evidence that the model is not dynamically complete. Later we will see an empirical illustration of this.

Now that a relation between the behaviour of the model errors and the specification of its systematic component was made clear, we are better equipped to handle the previous endogeneity problem. Which solutions are available to solve it? Many years ago the IV method was considered a good solution. However, it is quite obvious that this is not the case because although the inconsistency problem might be solved, the one of invalid inference methods is not because the errors remain autocorrelated. Then, a more reasonable solution is GMM, which generalizes the IV method and provides an asymptotically valid covariance matrix associated with it, supporting inference. Moreover, since autocorrelated errors may be dealt with in a framework of non-linear estimation, NLLS is also sometimes considered. However, both these solutions are far from satisfactory because the original problem is not really solved, this usually being one of insufficient dynamics. The same critic can

[5] The proof is simple when one starts from the alternative definition: the model is dynamically complete when $\mathrm{E}(u_t|\mathbf{x_t}, y_{t-1}, \mathbf{x_{t-1}}, \ldots) = 0, \forall t$.

be made, more generally, to the so called serial correlation corrections.

COMFAC. Indeed, none of the previous procedures can be generally recommended. As mentioned initially, residual autocorrelation symptoms are very often the tip of the iceberg of a misspecified regression function; usually they are not evidence of any evil intrinsic to the errors. More particularly, the British econometrics school, LSE-based and formerly lead by David Hendry (after Dennis Sargan), argues that serial correlation is often the result of a poor or insufficient dynamic specification, that the adoption of a general to specific (GTS) modeling strategy allows avoiding. The main idea is that we will be able to perform valid inferences only eliminating specification errors, and not "correcting" the estimation method with an AR(1) (or higher) model for the errors. As Grayham Mizon (1995) put it, "autocorrelation correctors: Don't!". Otherwise, if we opt for these corrections, reacting to autocorrelation symptoms augmenting the initial model with AR(1) errors, we may end up with a badly specified model and an inconsistent estimator.

To better understand this let us consider a somewhat exaggerated but common story where a practitioner begins by specifying the following static model, assuming also that its errors are serially uncorrelated:

$$\mathcal{M}_0 : \; y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t, \;\; u_t \text{ assumed } iid(0, \sigma_u^2),$$

where $\mathbf{x}_t$ now really contains only information which is contemporaneous to y_t. Therefore, since the model is static, it is very likely that (s)he finds serial correlation symptoms, which (s)he tends to correct dropping the initial i.i.d. assumption and adopting an AR(1) model for the errors: $u_t = \rho u_{t-1} + \epsilon_t$, $|\rho| < 1$, thereby estimating the transformed model where $\mathcal{M}_0$ is nested:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \rho(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta}) + \epsilon_t,$$

that is,

$$\mathcal{M}_1 : y_t - \rho y_{t-1} = (\mathbf{x}_t' - \rho\mathbf{x}_{t-1}')\boldsymbol{\beta} + \epsilon_t.$$

Now, denoting with $\boldsymbol{\gamma}$ the vector of coefficients of $\mathbf{x}_{t-1}$ in the unrestricted model, it must be noted that model $\mathcal{M}_1$ imposes the non-linear restrictions

$$\boldsymbol{\gamma} = -\boldsymbol{\rho}\,\boldsymbol{\beta}.$$

That is, the previous model is a restricted version of the more general model

$$\mathcal{M}_2 : y_t = \mathbf{x}_t'\boldsymbol{\beta} + \rho y_{t-1} + \mathbf{x}_{t-1}'\boldsymbol{\gamma} + \varepsilon_t, \; \varepsilon_t \sim iiid(0, \sigma^2),$$

where model $\mathcal{M}_1$ is nested, where those restrictions are not imposed. Model $\mathcal{M}_0$ is nested in $\mathcal{M}_1$ and $\mathcal{M}_1$ is nested in $\mathcal{M}_2$: $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \mathcal{M}_2$. Tests for serial correlation in $\mathcal{M}_0$ are tests for $\mathcal{M}_0$ against $\mathcal{M}_1$. Now, the rejection of the null (no serial correlation) with those tests may be due because the DGP is given by $\mathcal{M}_1$ but it may be also because $\mathcal{M}_0$ is misspecified (for instance, because the DGP is $\mathcal{M}_2$ or due to another unknown reason). Therefore, instead of "correcting" the serial correlation estimating $\mathcal{M}_1$, it is more reasonable to estimate first $\mathcal{M}_2$ and test the restrictions implied by $\mathcal{M}_1$ before "embarking" in the autocorrelation correction.

It is this test (or analysis) that is known as "common factor" test (or analysis) because we can write

$$\mathcal{M}_1 : (1-\rho L)y_t = (1-\rho L)\mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t,$$

$$\begin{aligned} \mathcal{M}_2 : (1-\rho L)y_t &= \mathbf{x}_t'\boldsymbol{\beta} + L\mathbf{x}_t'\boldsymbol{\gamma} + \varepsilon_t \\ &= \mathbf{x}_t'(\boldsymbol{\beta} + \boldsymbol{\gamma} L) + \varepsilon_t, \end{aligned}$$

where L denotes the usual lag operator ($L^p z_t = z_{t-p}$), i.e., in $\mathcal{M}_1$ but not in $\mathcal{M}_2$ the common factor $(1-\rho L)$ is present in both sides of the equation and this is the origin of the test's name. In relation to $\mathcal{M}_2$, $\mathcal{M}_1$ imposes the restriction of a common factor in both autoregressive polynomials of y_t and $\mathbf{x_t}$. Therefore, in this small world the departure point must be $\mathcal{M}_2$, the most general model, where those restrictions must be tested and only if they are considered valid by the data, that is, not rejected, we must proceed to estimate $\mathcal{M}_1$.

In conclusion, we must begin by specifying a general model and test restrictions on it so that it can be simplified, not the other way around. The first modelling strategy is known as "general to specific" (GTS) and, in logical terms, it appears to be clearly preferable to its inverse. In case the restrictions are not valid, the data not supporting them, this inverse strategy ends up with a misspecified model and with "inconsistent estimates".

Starting with the most general model, which in this simple case is $\mathcal{M}_2$, how to test the previous non-linear restrictions ($\boldsymbol{\gamma} = -\boldsymbol{\rho}\boldsymbol{\beta}$)? Since they are non-linear, the delta method[6] seems to be necessary. Actually, there are several approaches and Davidson and MacKinnon (1993, pp. 366-8) note that the simplest way to solve the problem is to resort to an asymptotic F-statistic. Notice also that this is a case where, contrarily to usual, the Wald testing principle is attractive because the

[6] See e.g. Hayashi (2000), pp. 93-4.

estimation of the general model, without restrictions, is easy because it is a linear model.

Note also that for this simplest AR(1) case, TSP provides automatically a test statistic. In fact, when any model with no lagged dependent variable as regressor is estimated in TSP and previous to the estimation instruction the option "`regopt (pvprint) auto;`" is used, TSP automatically calculates the Wald test for common factors and he calls the statistic "`Wald nonlin. AR1 vs. lags`". Symptoms of serial correlation together with the rejection of this hypothesis should lead us to re-specify the model introducing lags for all variables.

5 Testing for serial correlation

In this section the usual tests for error serial correlation are reviewed. These are the t-test for the case of strictly exogenous regressors, the Durbin-Watson test, Ljung-Box tests and the tests by Breusch and Godfrey. A basic idea, common to all the tests, is to replace the unobserved errors with the residuals.

5.1 The t-test for AR(1) errors with strict exogeneity

When the regressors are strictly exogenous a simple asymptotic t-test is available for AR(1) errors. The assumed model is $u_t = \rho u_{t-1} + \epsilon_t, |\rho| < 1$, and we wish to test $H_0 : \rho = 0$ usually versus $H_1 : \rho > 0$, though the alternative $H_1 : \rho \neq 0$ may be also considered.

To perform the test, the OLS residuals (e_t) are saved and they are used in the corresponding regression $e_t = \rho e_{t-1} + v_t$. The test statistic is $\widehat{\rho}/se(\widehat{\rho}) \overset{a}{\sim} \mathcal{N}(0,1)$ when H_0 is true. Correspondingly to the previous alternatives, an one-sided or two-sided critical region is used.

Though designed specifically for AR(1) errors, this test also has power to detect any other form of serial correlation provided that it is first order.

5.2 The Durbin-Watson statistic

With an ascendency as distinguished as the work of von Neumann, the Durbin-Watson statistic was the only available statistic to test for serial correlation in the linear regression model for many years. This is the reason why, although severely

limited, it is still automatically present in the standard regression output of many econometric packages.

The first of its limitations is the assumption of a stationary AR(1) model for the errors under the alternative: $u_t = \rho u_{t-1} + \epsilon_t$, $|\rho| < 1$. The second limitation is that its derivation was made under the assumption of deterministic regressors, which implies that strict exogeneity has to be assumed. In particular, it is known that its power can be very low to detect autocorrelation when y_{t-1} is a regressor. Actually, in that case the statistic becomes biased towards non-rejection even when there is serial correlation, clearly making the test powerless to detect the problem. Two other limitations are the assumptions of Gaussian errors and the presence of an intercept in the regression model.

The statistic is

$$DW = d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2},$$

and it is closely related with the OLS estimate of ρ in the regression $e_t = \rho e_{t-1} + v_t$ because it can be easily shown that

$$d \approx 2(1 - \widehat{\rho}).$$

Therefore, the domain of the statistic and its behavior can be easily derived from this approximate relation:

$$\begin{aligned} \text{if } \rho = 0 \Rightarrow \quad & \widehat{\rho} \approx 0 \Rightarrow \quad d \approx 2, \\ \text{if } \rho = 1 \Rightarrow \quad & \widehat{\rho} \approx 1 \Rightarrow \quad d \approx 0, \\ \text{if } \rho = -1 \Rightarrow \quad & \widehat{\rho} \approx -1 \Rightarrow \quad d \approx 4. \end{aligned}$$

Therefore, $d \in (0,\ 4)$ and the virtuous zone for the statistic is around 2. For macroeconomic time series one usually gets $d < 2$.

The major problem relates to the distribution of the statistic. In fact, recall that even when $\text{Var}(u|\mathbf{X}) = \sigma^2 I$ it can easily be shown that

$$\text{Var}(e|\mathbf{X}) = \sigma^2 \mathbf{M},$$

$\mathbf{M}$ denoting the usual anihilator matrix, $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Therefore:

a) even when the errors are serially uncorrelated the residuals will display some autocorrelation (since $\mathbf{M} \neq \mathbf{I}$), i. e., the OLS residuals do not totally reflect the properties of the errors.

b) Since $\mathbf{M} = f(\mathbf{X})$, it is impossible to obtain the exact distribution of the test statistic that does not depend on the particular $\mathbf{X}$ matrix, the observations of the regressors.

The solution provided by Durbin and Watson to overcome this problem was to derive the distributions of two statistics bounding from below and from above the distribution of d for any $\mathbf{X}$ matrix. These statistics are called d_L and d_U, respectively, and therefore, the decision rule is:

$$\begin{array}{lcl} \text{if } d < d_L & \Rightarrow & H_0 \text{ is rejected,} \\ \text{if } d > d_U & \Rightarrow & H_0 \text{ is not rejected, and} \\ \text{if } d_L < d < d_U & \Rightarrow & \text{the test is inconclusive.} \end{array}$$

This last case is obviously a major limitation of the test.

An informal solution to some of its problems is to use the statistic in an imprecise way, as a rough guide, for instance interpreting values below 1 or 1.1 or 1.2 as vaguely indicating some evidence for (positive) first order autocorrelation and totally disregarding any other values.

5.3 Ljung-Box tests

These tests are imported from the time series literature and will be briefly reviewed. Denote with $\rho_1, \rho_2, \ldots, \rho_j$, in general, the autocorrelation coefficients of the errors. These statistics aim to test

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_p = 0, \quad vs. \quad H_1 : \exists \rho_j \neq 0,\ j = 1, \ldots, p,$$

and for this reason they are called "portmanteau" test statistics: the purpose is to analyse the presence of serial correlation of all orders until a higher (than 1) order p. Denote with r_j the corresponding serial correlation coefficients of the OLS residuals. Then, a first statistic, the Box-Pierce statistic is

$$Q = T \sum_{j=1}^{p} r_j^2 \overset{a}{\sim} \chi^2_{(p)} \text{ under } H_0.$$

However, according to Ljung and Box, the quality of the approximation in small samples improves with the modified (LB) statistic

$$Q^* = T(T+2) \sum_{j=1}^{p} \frac{r_j^2}{T-j} \overset{a}{\sim} \chi^2_{(p)} \text{ under } H_0.$$

An obvious important problem, which is also common to the Breusch-Godfrey tests, is the choice of p. Anyway, notice again that the purpose is to test for the presence of autocorrelation *until* the p-th order, not only the p order. The choice must try to balance the two following arguments:

a) if p is (too) small, the test is unable to detect autocorrelation existing at a higher order;

b) if p is (too) large, the tests may lose power for a low order of autocorrelation because the significant autocorrelation coefficients are mixed with others that are non-significant and become diluted.

A first and quite obvious solution is to start chosing $p = 1$, precisely due to the reason that it is the order for which there are more tests available: when the dynamic specification is poor, it is the more recent past that is missing to capture the behaviour of the dependent variable; therefore, a test for first order autocorrelation is mandatory. Even clearly increasing the overall size beyond the usual 5%, a further test is usually performed, and its order depends on the data frequency. In case the data are annual, $p = 3$ (or even $p = 2$ only) is sufficient to gauge whether there are missing dynamics higher than the most recent one. For quarterly data, the model may not be able to capture the seasonal fluctuations of the dependent variable (assuming that they exist) and hence a test with $p = 4$ must be also performed. For the same reason, with monthly data the order $p = 12$ (and sometimes even $p = 6$ for inverse seasonality) is usually also selected. Moreover, in case the sample size is really large and one does not care whether the real overall size is large, $p = 8$ and $p = 24$ are also used sometimes for quarterly and monthly data, respectively.

In the time series literature, these tests are applied on the residuals of ARMA(m, q) models, and the distribution of the test statistics (under H_0) is $\chi^2_{(p-m-q)}$. In the regression framework the number of degrees of freedom is indeed p because no ARMA model is estimated on the OLS residuals. In this same framework, according to Godfrey (1998), the properties of these tests are not very well known, particularly when the regression includes a mixture of exogenous regressors and lagged dependent variable(s). Therefore, his recommendation is to use preferably the Breusch-Godfrey tests.

5.4 Breusch-Godfrey tests

These are the tests that should be used preferably because they don't have the limitations of the previous ones. Their origin is also straithgforward: they are Lagrange multiplier type tests that generalize Durbin's h-alt test for first order autocorrelation (that he proposed to replace the Durbin-Watson statistic in the case where the lagged dependent variable is one of the regressors).

For this case, the test statistic is the t-ratio for ϕ from the auxiliar regression

$$e_t = \mathbf{x}_t'\boldsymbol{\alpha} + \phi e_{t-1} + v_t,$$

which, under H_0, is asymptotically distributed as $\mathcal{N}(0,1)$. Actually, this regression should be performed not only in that case but in any other case where x_t contain any stochastic regressor. Since very often at least one of these regressors is not strictly exogenous, the OLS residuals are not good estimates of the model errors. The repetead presence of the regressors in the auxiliar test regression has the purpose to clean the residuals (both current and lagged) of any behaviour due to the (non-contemporaneous) correlation between the errors and the regressors, so that the properties of the errors may become more visible.

In the general case, the null and alternative hypothesis of the Breusch-Godfrey tests are the same as the Ljung-Box ones. However, they can also be put in a different and attractive framework:

$$H_0 : u_t \text{ not autocorrelated } \; vs. \; H_1 : u_t \text{ is } ARMA(m,q), \forall m, q : m + q = p.$$

There are several alternative ways to derive the test statistics. As they are based on the Lagrange multiplier principle (see, e.g. Johnston e DiNardo, 1997, or Martin et al. 2013), they are simple to perform because only the model under the null hypothesis (without autocorrelation) needs to be estimated. Martin et al. (2013) contain several alternative derivations for the simple case of a test against the AR(1) alternative hypothesis, where all the calculations are presented. A more general and elegant derivation is presented in Davidson and MacKinnon (1993) for the approach based on the artificial Gauss-Newton regressions. This approach also allows us to understand that the test statistic against AR(1) errors is the same as the test statistic against MA(1) errors; as Davidson and MacKinnon (1993, p. 359) emphasize, they are locally equivalent alternatives (that is, models that have identical derivatives when evaluated under the null hypothesis). Moreover, the test statistic against

AR(p) errors is the same as the one for testing against MA(p) errors, as well as the same as the one for ARMA(m, q) errors, with $m + q = p$. [7] In Hayashi (2000) the derivation is made through the asymptotic equivalence to a version of the Box-Pierce statistic.

The test procedure is a simple extension of the h-alt one:

i) perform the regression of the OLS residuals on all the regressors of the model and on the lagged residuals $e_{t-1}, e_{t-2}, \ldots, e_{t-p}$, where the first missing residuals may be replaced with zeros,

$$e_t = \mathbf{x}_t'\boldsymbol{\alpha} + \phi_1 e_{t-1} + \ldots + \phi_p e_{t-p} + error_t,$$

that is, $\mathbf{e} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{E}\boldsymbol{\phi} + \mathbf{error}$, $\mathbf{E}$ denoting the matrix of the lagged residuals;

ii) test

$$H_0 : \phi_1 = \phi_2 = \ldots = \phi_p = 0, \;\; vs. \;\; H_1 : \exists \phi_j \neq 0, j = 1, \ldots, p,$$

with the usual (LM) test statistic

$$BG(p) = TR_e^2 \overset{a}{\sim} \chi^2_{(p)} \text{ under } H_0,$$

where R_e^2 denotes the R-squared of the auxiliar regression.

Alternatively, the (asymptotic) F statistic for this same test may be used and it may possess better small sample properties. As Greene (2012) notes, the test regression is equivalent to the regression of the residuals on the part of the lagged residuals that is not explained by $\mathbf{X}$ (see exercise 6).

6 Empirical example

More than to to exemplify the use of test statistics for serial correlation, the following example illustrates the emergence of residual autocorrelation symptoms as resulting from a poorly dynamically specified model. The context is the portuguese economy, the data are quarterly and the sample is from 1977:1 to 1995:4. The following model was estimated

$$LMR_t = \beta_1 + \delta_1 Q_{t1} + \delta_2 Q_{t2} + \delta_2 Q_{t3} + \beta_2 t + \beta_2 LGDP_t + \beta_4 r_t + u_t, \;\; t = 1, 2, \ldots, T,$$

[7] This is the reason for the title adopted in TSP for these test statistics; see below, the section on the empirical example.

LMR denoting the logarithm of real M1 (physical currency in circulation and demand deposits but not time deposits), $Q_{tj}, j = 1, \ldots, 4$ are the usual quarterly dummies, LGP represents the logarithm of GDP and r is an interest rate for time deposits. Therefore, altough containing a time variable, the model is purely static. (Some of) The *output* provided by TSP is the folowing:

```
                                        Equation   1
                                        ============
                      Method of estimation = Ordinary Least Squares
Dependent variable: LMR
Current sample:  1977:1 to 1995:4
         Mean of dep. var. = 8.94461
  Std. error of regression = .044500
                 R-squared = .936480
              LM het. test = .486913 [.485]
             Durbin-Watson = .846703 ** [.000,.000]
Breusch/Godfrey LM: AR/MA1 = 33.5133 ** [.000]
Breusch/Godfrey LM: AR/MA4 = 34.6787 ** [.000]
 Wald nonlin. AR1 vs. lags = 16.1423 ** [.000]
           Ramsey's RESET2 = .837736 [.363]
           F (zero slopes) = 169.546 [.000]
            Schwarz B.I.C. = -117.207
           Estimated    Standard
Variable  Coefficient     Error       t-statistic   P-value
C         1.82332       1.08542       1.67983       [.098]
Q1        -.072107      .014877       -4.84672      [.000]
Q2        -.068193      .014631       -4.66077      [.000]
Q3        -.049859      .014505       -3.43744      [.001]
T         -.649085E-02  .103312E-02   -6.28277      [.000]
LGDP      1.02708       .146509       7.01035       [.000]
R         -2.46789      .112671       -21.9036      [.000]
```

Although with a very good fit to the data, with all regressors highly significant and easily passing an heteroskedasticity test and a RESET test for functional form (with only the squares of the fitted values), the model fails completely as regards the hypothesis of no serial correlation for the errors, which is hardly a surprise. Besides a

DW statistic (employed informally and loosely) somewhat below 1, both the BG(1) and BG(4) statistics are highly significant, both with p-values equal to 0.000, and hence very strongly allowing the rejection of the null. This strong evidence for the presence of serial correlation is the only, but very significant, major shortcoming of the model. The 5% critical regions for these tests are, respectively,

$$CR_{BG(1)} = \{BG(1) : BG(1) > 3.841\} \text{ and } CR_{BG(4)} = \{BG(4) : BG(4) > 9.488\}.$$

Note also that, generally speaking, the significant BG(4) may be due not only to the presence of first order serial correlation but also to a failure in explaining seasonality, that is, the presence of the seasonal dummies does not guarantee that seasonality in LMR is totally captured by them; there may be stochastic seasonality that is not captured by the seasonal dummies. Since BG(4) is only a bit higher than BG(1) it appears that this is not the case here. Anyway, it could be significant also due to the presence of second or third order serial correlation.

Notice, moreover, that the `Wald nonlin. AR1 vs. lags` statistic clearly rejects the possibility that the first order autocorrelation problem can be solved with a "correction" estimation method for serial correlation. (Which are the null and alternative hypothesis here?)

To illustrate that the autocorrelation problems are seemingly reflecting only a poor dynamic specification, a slight (but often powerful) modification was made: the introduction of the lagged dependent variable as a regressor (LMR_{t-1}). Notice that this introduction does not aim to serve as a general method. Instead, as will become clear later, one should begin with a much more general model and test down until a reasonable and data coherent specification is reached. The purpose here is to illustrate that symptoms of residual autocorrelation arise not as an error problem but, instead, from missing dynamics.

The new estimated equation is presented below. Observe that LMR_{t-1} appears as highly significant but far away from 1 (the importance of this will become clear further ahead, in the next chapter). Notice also that TSP now automatically prints the h-alt statistic, and it is very far from significant, with an asymptotic p-value that is almost 1. Obviously, it equals the p-value of the BG(1) statistic, as expected. Moreover, the BG(4) statistic is also clearly insignificant, with a p-value exceding 0.80. Obviously, the critical regions of both tests are the same of the previous model.

Notice that we are not certain that the serial correlations problems really result from the insufficient dynamic specification. But the evidence supporting that claim

is indeed very strong.

```
                                        Equation   2
                                        ============
                      Method of estimation = Ordinary Least Squares
Dependent variable: LMR
Current sample:  1977:2 to 1995:4

  Sum of squared residuals = .051269
     Variance of residuals = .765202E-03
  Std. error of regression = .027662
                 R-squared = .976019
              LM het. test = .251423 [.616]
            Durbin-Watson = 1.95318 [.132,.760]
           Durbin's h alt. = -.054748 [.956]
Breusch/Godfrey LM: AR/MA1 = .299737E-02 [.956]
Breusch/Godfrey LM: AR/MA4 = 1.45031 [.835]
           Ramsey's RESET2 = .403427 [.528]
           F (zero slopes) = 389.549 [.000]
            Schwarz B.I.C. = -149.616


            Estimated    Standard
Variable  Coefficient     Error       t-statistic   P-value
C         1.16358       .681946       1.70626       [.093]
Q1        -.119098      .010311       -11.5504      [.000]
Q2        -.058601      .914345E-02   -6.40912      [.000]
Q3        -.038966      .907600E-02   -4.29334      [.000]
T         -.177309E-02  .782624E-03   -2.26557      [.027]
LMR(-1)   .623910       .059107       10.5556       [.000]
LGDP      .325642       .113377       2.87221       [.005]
R         -.954095      .160583       -5.94144      [.000]
```

7 Appendix

7.1 Parametric estimation of the long-run variance

Let us suppose that y_t follows a stationary ARMA(p, q) process, $\phi(L)y_t = \theta(L)\epsilon_t$. Then, its spectrum is given by

$$S(\omega) = \frac{1}{2\pi}\sigma_\epsilon^2 \frac{\theta(e^{-i\omega})\theta(e^{i\omega})}{\phi(e^{-i\omega})\phi(e^{i\omega})},$$

where i denotes the imaginary unit ($i^2 = -1$). Therefore,

$$\begin{aligned} S &= 2\pi S(0) \\ &= \sigma_\epsilon^2 \tfrac{\theta(1)^2}{\phi(1)^2} \\ &= \sigma_\epsilon^2 \tfrac{(1+\theta_1+\theta_2+\ldots+\theta_q)^2}{(1-\phi_1-\phi_2-\ldots-\phi_q)^2}. \end{aligned}$$

Hence, to estimate the long-run variance of the process we only need to estimate the ARMA model and subsequently use

$$\widehat{S} = \sigma_\epsilon^2 \frac{(1+\widehat{\theta}_1+\widehat{\theta}_2+\ldots+\widehat{\theta}_p)^2}{(1-\widehat{\phi}_1-\widehat{\phi}_2-\ldots-\widehat{\phi}_p)^2}.$$

7.2 COMFAC in higher order AR models

The common factor restrictions are also implicit in linear regression models with autoregressive errors of any order. For instance, if the model is estimated assuming an AR(2) process for the errors,

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \epsilon_t,$$

then it can be written as

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = (\mathbf{x}_t' - \phi_1 \mathbf{x}_{t-1}' - \phi_2 \mathbf{x}_{t-2}')\boldsymbol{\beta} + \epsilon_t,$$

that is

$$(1 - \phi_1 L - \phi_2 L^2)y_t = (1 - \phi_1 L - \phi_2 L^2)\mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t$$

which is the restricted version of

$$(1 - \phi_1 L - \phi_2 L^2)y_t = \mathbf{x}_t'\boldsymbol{\beta} + L\mathbf{x}_t'\boldsymbol{\gamma}_1 + L^2\mathbf{x}_t'\boldsymbol{\gamma}_2 + \epsilon_t.$$

Therefore, in the second equation the factor $(1 - \phi_1 L - \phi_2 L^2)$ is common to both autoregressive polynomials, a situation that does not occur with the last equation. The tests for the common factors are very similar to those of the models with AR(1) errors. A discussion of these tests appears in Davidson and MacKinnon (1993).

8 Additional readings

The following texts are also recommended.

- At an introductory level: chapter 12 of Wooldridge (2016) contains a classical approach and addresses the correction approach with FGLS as well.
- At a more advanced level Spanos (2018) offers a comprehensive and recent perspective on misspecification testing. Godfrey (1988) is a classical and solid reading on this subject. West (2010) is a very condensed and somewhat recent update on the topic of HAC estimation.

Bibliography

Davidson, R. and MacKinnon, J. G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

Davidson, R. and MacKinnon, J. G. (2004), *Econometric Theory and Methods*, Oxford University Press.

Godfrey, L. G. (1988), *Misspecification tests in Econometrics*, Cambridge University Press.

Greene, W. H. (2012), *Econometric Analysis*, 7th ed., Pearson Education.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

Hayashi, F. (2000), *Econometrics*, Princeton University Press, Princeton.

Johnston, J. and DiNardo, J. (1997), *Econometric Methods*, McGraw-Hill.

Martin, V., Hurn, S. and Harris, D. (2013), *Econometric Modelling with Time Series*, Cambridge University Press.

Mizon, G. (1995), A simple message for autocorrelation correctors: Don't, *Journal of Econometrics*, 69 (1), pp. 267-83.

Newey, W. K. and and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55 (3), pp. 703-708.

Spanos, A. (2018), Mis-specification testing in retrospect, *Journal of Economic Surveys*, 32 (2), pp. 541-77.

Stewart, J. (1991), *Econometrics*, Prentice Hall.

Stock, J. H. and Watson, M. W. (2015), *Introduction to Econometrics*, 3rd ed., Pearson Addison-Wesley.

West, K. (2010), Heteroskedasticity and autocorrelation corrections, *in* Durlauf, S. N. and Blume, L. E. (eds.), *Macroeconometrics and Time Series Analysis*, Palgrave MacMillan.

Wooldridge, J. M. (2016), *Introductory Econometrics, a Modern Approach*, 6th ed., Cengage Learning.

Exercises

1. (From a previous exam.) Explain what is and what is the usefulness of "HAC" estimation. Which conditions are necessary for its valid use?

2. (Adapted from an old exam.) State, whether the following declaration is true or false: assuming that we wish to analyse the dynamic effects of x_t over y_t, if the errors of the regression

$$y_t = \beta_1 + \beta_2 \, x_t + u_t$$

   show symptoms of serial correlation, we should use the Newey-West matrix, that is, an "HAC" estimator for the covariance matrix of the OLS estimator.

3. Consider the model $y_t = \beta_1 + \beta_2 x_t + u_t$, with $u_t = \rho u_{t-1} + \epsilon_t$, $|\rho| < 1$. Prove that it is not dynamically complete.

4. Consider the Durbin-Watson statistic:

$$\text{DW} = d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}.$$

Show that $d \approx 2(1 - \widehat{\rho})$, where $\widehat{\rho}$ is the OLS estimator from the regression of e_t on e_{t-1} (with no intercept).

5. (Taken from an exam of the course for undergraduates.) Assuming that $\{(y_t, x_t)\}$ is stationary and weakly dependent, the next models were estimated with 50 annual observations:

(A)
$$\begin{array}{lllll} y_t = & 13.88 + & 0.796\, x_t + & 1.608\, x_{t-1} + \hat{u}_t, & \hat{\hat{u}}_t = \quad 0.555\, \hat{u}_{t-1} \\ & (1.58) & (0.235) & (0.233) & \qquad\qquad (0.124) \end{array}$$

(B)
$$\begin{array}{llllll} y_t = & 3.26 + & 0.700\, x_t + & 0.470\, x_{t-1} + & 0.604\, y_{t-1} + \hat{v}_t, & \\ & (1.39) & (0.134) & (0.175) & (0.061) & \end{array}$$

$$\begin{array}{lllllll} \hat{\hat{v}}_t = & 0.063 + & 0.006\, x_t + & 0.015\, x_{t-1} - & 0.007\, y_{t-1} + & 0.031\, \hat{v}_{t-1} \\ & (1.50) & (0.138) & (0.188) & (0.067) & (0.164) \end{array}$$

   a) Providing a formal answer, perform the statistical tests ($\alpha = 0.05$) that resort to the auxiliar regressions of both models and draw the most appropriate conclusion from each of them.

   b) Which is the global conclusion that you can draw from the comparison between the results of the previous tests? Justify your answer.

6. Show that the regressions from the Breusch-Godfrey tests are equivalent to those from e_t (the OLS residuals) "over the part of the lagged residuals that is not explained by $\mathbf{X}$" (Greene, 2012, p. 962).

7. Now it is time to do 1.b) and 1.c) of the previous set of exercises.